1. A method for fabricating a MONOS memory device comprising:

forming a gate silicon oxide layer on the surface of a semiconductor substrate;

depositing a first polysilicon layer overlying said gate silicon oxide layer;

depositing a first nitride layer overlying said first polysilicon layer;

patterning said first polysilicon layer and said first nitride layer to form word gates wherein a gap is left between two of said word gates;

forming a first insulating layer on the sidewalls of said word gates;

depositing a spacer layer overlying said word gates and said gate silicon oxide layer;

anisotropically etching away said spacer layer to leave disposable spacers on the sidewalls of said word gates;

implanting ions into said semiconductor substrate to form a lightly doped region wherein said disposable spacers act as an implantation mask;

thereafter removing said disposable spacers;

depositing a nitride-containing layer over said semiconductor substrate in said gap;

depositing a second polysilicon layer overlying said word gates and said nitride-containing layer;

anisotropically etching away said second polysilicon layer and said nitride-containing layer to leave polysilicon spacers on the sidewalls of said word gates wherein said polysilicon spacers form control sidewall spacer gates and wherein said nitride-containing layer underlying each of said control sidewall spacer gates forms a nitride region in which charge is stored ;

forming a second insulating layer on said control sidewall spacer gates;

implanting ions into said semiconductor substrate to form a bit diffusion region wherein said control sidewall spacer gates act as an implantation mask;

coating a gap filling material over the surface of said substrate wherein said gap-filling material fills said gap between said two of said word gates;

planarizing said gap-filling material;

thereafter removing said first nitride layer overlying said word gates; and

depositing a third polysilicon layer overlying said substrate wherein said third polysilicon layer forms a word line connecting underlying said word gates to complete said fabrication of said MONOS memory device.

2. The method according to Claim 1 wherein said gate silicon oxide layer has a thickness of between about 5 and 10 nanometers.

3. The method according to Claim 1 wherein said first polysilicon layer is deposited by chemical vapor deposition to a thickness of between about 150 and 250 nanometers.

4. The method according to Claim 1 wherein said first nitride layer is deposited by chemical vapor deposition to a thickness of between about 50 and 100 nanometers.

5. The method according to Claim 1 wherein said first insulating layer is formed by thermally growing a silicon oxide layer to a thickness of between about 5 and 10 nanometers on the sidewalls of said word gates.

6 The method according to Claim 1 wherein said first insulating layer is formed by depositing a silicon oxide layer by chemical vapor deposition to a thickness of between about 5 and 10 nanometers on the sidewalls of said word gates.

7. The method according to Claim 1 wherein said first insulating layer is formed by depositing a silicon nitride layer to a thickness of between about 5 and 10 nanometers on the sidewalls of said word gates.

8. The method according to Claim 1 wherein said first insulating layer is formed by depositing a silicon oxide layer and a silicon nitride layer to a combined thickness of between about 5 and 10 nanometers on the sidewalls of said word gates.

9. The method according to Claim 1 wherein said spacer layer comprises one of the group containing polysilicon, plasma nitride, plasma oxynitride, and borophosphosilicate glass (BPSG) and has a thickness of between about 30 and 50 nanometers.

10. The method according to Claim 1 wherein said step of removing said disposable spacers comprises a dry chemical anisotropic etch.

11. The method according to Claim 1 wherein said step of depositing said nitride-containing layer comprises:

growing a first silicon oxide layer to a thickness of between about 3.6 and 5.0 nanometers on said semiconductor substrate;

depositing a silicon nitride layer having a thickness of about 2 to 5 nanometers overlying said first silicon oxide layer; and

depositing a second silicon oxide layer having a thickness of between about 4 and 8 nanometers overlying said silicon nitride layer.

12. The method according to Claim 1 further comprising nitriding said first silicon oxide layer before said step of depositing said silicon nitride layer.

13. The method according to Claim 1 wherein said second polysilicon layer has a thickness of between about 30 and 50 nanometers.

14. The method according to Claim 1 wherein said second polysilicon layer has a thickness of between about 30 and 50 nanometers and further comprising depositing a tungsten silicide layer having a thickness of between about 60 and 100 nanometers and wherein said second polysilicon layer and tungsten silicide layer together form said control sidewall spacer gates.

15. The method according to Claim 1 wherein said second insulating layer comprises silicon oxide deposited by chemical vapor deposition to a thickness of about 10 nanometers.

16. The method according to Claim 1 wherein said second insulating layer comprises silicon nitride deposited by chemical vapor deposition to a thickness of about 10 nanometers.

17. The method according to Claim 1 further comprising:

anisotropically etching said second insulating layer to form sidewall oxide spacers on lower portions of said control sidewall spacer gates; and

thereafter siliciding upper portions of said control sidewall spacer gates and said bit diffusion region.

18. The method according to Claim 1 wherein said gap-filling material comprises one of the group containing silicon oxide and borosilicate glass.

19. The method according to Claim 1 wherein said gap-filling material comprises a conductive material and further comprising:

recessing said conductive material below the surface of said first nitride layer;

depositing a silicon oxide layer overlying said recessed conductive material; and

planarizing said silicon oxide layer wherein said conductive material and underlying said control sidewall spacer gates together form a control gate.

20. The method according to Claim 1 wherein said third polysilicon layer has a thickness of between about 150 and 200 nanometers.

21. The method according to Claim 1 further comprising siliciding said word line.

22. A method for fabricating a step split structure MONOS memory device comprising:

forming a gate silicon oxide layer on the surface of a semiconductor substrate;

depositing a first polysilicon layer overlying said gate silicon oxide layer;

depositing a first nitride layer overlying said first polysilicon layer;

patterning said first polysilicon layer and said first nitride layer to form word gates wherein a gap is left between two of said word gates;

forming a first insulating layer on the sidewalls of said word gates;

depositing a spacer layer overlying said word gates and said gate silicon oxide layer;

anisotropically etching away said spacer layer to leave disposable spacers on the sidewalls of said word gates;

etching away said gate silicon oxide layer not covered by said word gates and said disposable spacers to expose a portion of said semiconductor substrate;

etching away said exposed portion of said semiconductor substrate to form a step into said substrate;

implanting ions into said semiconductor substrate to form a lightly doped region wherein said disposable spacers act as an implantation mask;

thereafter removing said disposable spacers;

removing said gate silicon oxide layer underlying said disposable polysilicon spacers;

forming a composite layer of oxide-nitride-oxide overlying said semiconductor substrate;

depositing a second polysilicon layer overlying said word gates and said second gate silicon oxide layer;

anisotropically etching away said second polysilicon layer and said composite oxide-nitride-oxide layer to leave polysilicon spacers on the sidewalls of said word gates wherein said polysilicon spacers form sidewall control gates and wherein the nitride portion of said composite oxide-nitride-oxide layer underlying each of said sidewall control gates forms a nitride region in which charge is stored;

forming a second insulating layer on said control sidewall gates;

implanting ions into said semiconductor substrate to form a bit diffusion region wherein said control sidewall gates act as an implantation mask;

coating a gap filling material over the surface of said substrate wherein said gap-filling material fills said gap between said two of said word gates;

planarizing said gap-filling material;

thereafter removing said first nitride layer overlying said word gates; and

depositing a third polysilicon layer overlying said substrate wherein said third polysilicon layer forms a word line connecting underlying said word gates to complete said fabrication of said MONOS memory device.

23. The method according to Claim 22 wherein said first polysilicon layer is deposited by chemical vapor deposition to a thickness of between about 150 and 250 nanometers.

24. The method according to Claim 22 wherein said first nitride layer is deposited by chemical vapor deposition to a thickness of between about 50 and 100 nanometers.

25. The method according to Claim 22 wherein said first insulating layer is formed by thermally growing a silicon oxide layer to a thickness of between about 5 and 10 nanometers on the sidewalls of said word gates.

26. The method according to Claim 22 wherein said first insulating layer has a thickness of between about 5 and 10 nanometers on the sidewalls of said word gates.

27. The method according to Claim 22 wherein said spacer layer comprises one of the group containing polysilicon, plasma nitride, plasma oxynitride, and borophosphosilicate glass (BPSG) and has a thickness of between about 30 and 50 nanometers.

28. The method according to Claim 22 wherein said step of removing said disposable spacers comprises a dry chemical anisotropic etch.

29. The method according to Claim 22 wherein said step into said semiconductor substrate has a depth of between about 20 and 50 nanometers.

30. The method according to Claim 22 after said step of removing said gate silicon oxide layer underlying said disposable spacers further comprising rounding the corners of said step.

31. The method according to Claim 30 wherein said step of rounding said corners of said step comprises a rapid thermal anneal at between about 1000 and 1100 °C for about 60 seconds.

32. The method according to Claim 30 wherein said step of rounding said corners of said step comprises annealing in hydrogen at about 900 °C at a pressure of between about 200 and 300 mtorr.

33. The method according to Claim 22 wherein said oxide-nitride-oxide composite layer comprises:

a first silicon oxide layer having a thickness of between about 3.6 and 5.0 nanometers;

a second silicon nitride layer having a thickness of about 2 to 5 nanometers; and

a third silicon oxide layer having a thickness of between about 4 and 8 nanometers.

34. The method according to Claim 22 wherein said second polysilicon layer has a thickness of between about 30 and 50 nanometers.

35. The method according to Claim 22 wherein said second polysilicon layer has a thickness of between about 30 and 50 nanometers and further comprising depositing a tungsten silicide layer having a thickness of between about 60 and 100 nanometers and

wherein said third polysilicon layer and tungsten silicide layer together form said control sidewall spacer gates.

36. The method according to Claim 22 wherein said second insulating layer comprises silicon oxide deposited by chemical vapor deposition to a thickness of about 10 nanometers.

37. The method according to Claim 22 wherein said second insulating layer comprises silicon nitride deposited by chemical vapor deposition to a thickness of about 10 nanometers.

38. The method according to Claim 22 further comprising:

anisotropically etching said second insulating layer to form sidewall oxide spacers on lower portions of said control sidewall spacer gates; and

thereafter siliciding upper portions of said control sidewall spacer gates and said bit diffusion region.

39. The method according to Claim 22 wherein said gap-filling material comprises one of the group containing silicon oxide and borosilicate glass.

40. The method according to Claim 22 wherein said gap-filling material comprises a conductive material and further comprising:

recessing said conductive material below the surface of said first nitride layer;

depositing a silicon oxide layer overlying said recessed conductive material; and

planarizing said silicon oxide layer wherein said conductive material and underlying said control sidewall spacer gates together form a control gate.

41. The method according to Claim 22 wherein said third polysilicon layer has a thickness of between about 90 and 180 nanometers.

42. The method according to Claim 22 further comprising siliciding said word line.

43. The method according to Claim 22 further comprising siliciding said word line.

44. A method for fabricating a MONOS memory device comprising:

forming a nitride-containing layer on the surface of a semiconductor substrate;

depositing a first polysilicon layer overlying said nitride-containing layer;

depositing a second nitride layer overlying said first polysilicon layer;

patterning said first polysilicon layer and said second nitride layer to form word gates wherein a gap is left between two of said word gates;

forming a first insulating layer on the sidewalls of said word gates;

depositing a spacer layer overlying said word gates and said gate silicon oxide layer;

anisotropically etching away said spacer layer to leave disposable spacers on the sidewalls of said word gates;

implanting ions into said semiconductor substrate to form a bit diffusion junction wherein said disposable spacers act as an implantation mask;

thereafter removing said disposable spacers;

depositing a second polysilicon layer overlying said word gates and filling said gap;

recessing said second polysilicon layer below a surface of said second nitride layer;

siliciding said recessed second polysilicon layer wherein said silicided recessed second polysilicon layer forms a control gate;

depositing an oxide layer overlying said silicided recessed second polysilicon layer;

thereafter removing said second nitride layer overlying said word gates; and

depositing a third polysilicon layer overlying said substrate wherein said third polysilicon layer forms a word line connecting underlying said word gates to complete said fabrication of said MONOS memory device.

45. The method according to Claim 44 wherein said step of forming said nitride-containing layer comprises:

growing a first silicon oxide layer to a thickness of between about 3.6 and 5.0 nanometers on said semiconductor substrate;

depositing a silicon nitride layer having a thickness of about 2 to 5 nanometers overlying said first silicon oxide layer; and

depositing a second silicon oxide layer having a thickness of between about 4 and 8 nanometers overlying said silicon nitride layer.

46. The method according to Claim 45 further comprising nitriding said first silicon oxide layer before said step of depositing said silicon nitride layer.

47. The method according to Claim 44 wherein said first polysilicon layer is deposited by chemical vapor deposition to a thickness of between about 150 and 250 nanometers.

48. The method according to Claim 44 wherein said second nitride layer is deposited by chemical vapor deposition to a thickness of between about 50 and 100 nanometers.

49. The method according to Claim 44 wherein said first insulating layer is formed to a thickness of between about 5 and 10 nanometers on the sidewalls of said word gates.

50. The method according to Claim 44 wherein said spacer layer comprises one of the group containing polysilicon, plasma nitride, plasma oxynitride, and borophosphosilicate glass (BPSG) and has a thickness of between about 30 and 50 nanometers.

51. The method according to Claim 44 further comprising before said step of removing said disposable spacers:

etching away said second silicon oxide layer not covered by said disposable spacers;

depositing a third silicon oxide layer overlying said nitride layer to a thickness of between about 4 and 6 nanometers; and

oxidizing said third silicon oxide layer to form an oxide layer having a thickness of about 20 nanometers over said nitride layer whereby coupling capacitance between said control gate and said bit diffusion is reduced.

52. The method according to Claim 44 wherein said step of removing said disposable spacers comprises a dry chemical anisotropic etch.

53. The method according to Claim 44 wherein said second polysilicon layer has a thickness of between about 30 and 50 nanometers.

54. The method according to Claim 44 wherein said second insulating layer comprises silicon oxide deposited by chemical vapor deposition to a thickness of about 10 nanometers.

55. The method according to Claim 44 wherein said second insulating layer comprises silicon nitride deposited by chemical vapor deposition to a thickness of about 10 nanometers.

56. The method according to Claim 44 wherein said third polysilicon layer has a thickness of between about 150 and 200 nanometers.

57. A method for fabricating a flash memory device comprising:

providing word gates overlying a gate silicon oxide layer on the surface of a semiconductor substrate wherein a gap is left between two of said word gates;

forming disposable spacers on the sidewalls of said word gates;

implanting ions into said semiconductor substrate to form a lightly doped region wherein said disposable spacers act as an implantation mask;

thereafter removing said disposable spacers;

forming sidewall polysilicon gates on the sidewalls of said word gates, each of said sidewall polysilicon gates having an underlying nitride-containing layer wherein the nitride region of said nitride-containing layer acts as a nitride charge region;

implanting ions into said semiconductor substrate to form a bit diffusion region wherein said sidewall polysilicon gates act as an implantation mask;

forming an insulating layer on said sidewall gates;

filling said gap between said two of said word gates with a second polysilicon layer;

recessing said second polysilicon layer;

siliciding said recessed second polysilicon layer;

covering said silicided recessed second polysilicon layer with an oxide layer wherein said silicided recessed second polysilicon layer along with underlying said sidewall polysilicon gates form a control gate; and

depositing a third polysilicon layer overlying said substrate wherein said third polysilicon layer forms a word line connecting said word gates to complete said fabrication of said flash memory device.

58. The method according to Claim 57 wherein said first polysilicon layer has a thickness of between about 150 and 250 nanometers.

59. The method according to Claim 57 wherein said disposable spacers comprise one of the group containing polysilicon, plasma nitride, plasma oxynitride, and borophosphosilicate glass (BPSG).

60. The method according to Claim 57 wherein said nitride-containing layer comprises a first layer of silicon oxide, a second layer of silicon nitride, and a third layer of silicon oxide.

61. The method according to Claim 57 after said step of removing said disposable spacers further comprising etching into said semiconductor substrate to form a step into said semiconductor substrate having a depth of between about 20 and 50 nanometers.

62. The method according to Claim 57 further comprising rounding the corners of said step.

63. The method according to Claim 62 wherein said step of rounding said corners of said step comprises a rapid thermal anneal at between about 1000 and 1100 °C for about 60 seconds.

64. The method according to Claim 62 wherein said step of rounding said corners of said step comprises annealing in hydrogen at about 900 °C at a pressure of between about 200 and 300 mtorr.

65. The method according to Claim 57 wherein a channel length defined from an edge of said word gate to an edge of adjacent said bit diffusion region is between about 30 and 50 nm and whereby ballistic electron injection occurs.

66. A MONOS memory cell comprising:

a word gate on the surface of a semiconductor substrate;

sidewall control gates on sidewalls of said word gate, separated from said word gates by an insulating layer;

nitride regions within an ONO layer underlying said sidewall control gates wherein electron memory storage is performed within said nitride regions;

a polysilicon word line overlying and connecting said word gate with word gates in other said memory cells and overlying said sidewall control gates, separated from said sidewall control gates by an insulating layer; and

bit line diffusions within said semiconductor substrate adjacent to each of said sidewall control gates.

67. The MONOS memory cell of Claim 66 wherein each sidewall control gate is separated from a sidewall control gate of another said memory cell by an insulating layer.

68. The MONOS memory cell of Claim 66 wherein each control gate comprises a polysilicon layer between two of said word gates overlying said bit diffusion region and said sidewall control gates wherein said nitride regions underlie only said sidewall control gates.

69. The MONOS memory call of Claim 66 wherein a channel length defined from an edge of said word gate to an edge of adjacent said bit diffusion region is between about 30 and 50 nm and whereby ballistic electron injection occurs.

70. The MONOS memory cell of Claim 66, wherein one of said nitride regions is a selected nitride region, and the other of said nitride regions is an unselected nitride region, and wherein said bit line diffusion near said selected nitride region is a bit diffusion, and said bit line diffusion near said unselected nitride region is a source diffusion, wherein a read operation of said cell is performed by:

over-riding said unselected nitride region;

providing a voltage on said word gate having a sum of the word gate threshold voltage, an overdrive voltage, and the voltage on said source diffusion;

providing a voltage on said control gate adjacent to said selected nitride region sufficient to allow for reading of the selected nitride region; and

reading said cell by measuring the voltage level on said bit diffusion.

71. The MONOS memory cell of Claim 70 wherein said memory cell is one of many cells in a MONOS memory array, and further comprising applying a control gate voltage of 0 volts to all cells beside the cell desired to be read.

72. The MONOS memory cell of Claim 70 wherein said memory cell is one of many cells in a MONOS memory array, and further comprising applying a control gate voltage of –0.7 volts to all cells beside the cell desired to be read in order to stop leakage.

73. The MONOS memory cell of Claim 66 wherein the voltage level on said bit diffusion may represent one of multiple threshold levels of said cell.

74. The MONOS memory cell of Claim 66, wherein one of said nitride regions is a selected nitride region, and the other of said nitride regions is an unselected nitride region, and wherein said bit line diffusion near said selected nitride region is a bit diffusion, and said bit line diffusion near said unselected nitride region is a source diffusion, wherein a program operation of said cell is performed by:

providing a high voltage on said unselected control gate to over-ride said unselected nitride region;

raising the control gate voltage of said selected nitride region;

providing a fixed voltage on said bit diffusion;

providing a voltage on said word line which is greater than said word gate threshold voltage; and

lowering the voltage of said source diffusion such that current flows from said source diffusion to said bit diffusion wherein ballistic injection of electrons occurs from a channel region to said selected nitride region when current flows.

75. The MONOS memory cell of Claim 74 wherein multiple thresholds can be programmed by varying said voltage on said bit diffusion line.

76. The MONOS memory cell of Claim 74 wherein said memory cell is one of many cells in a MONOS memory array, and further comprising disabling nitride regions in adjacent cells sharing a word line by applying a control gate voltage of 0 volts to said adjacent cells.

77. The MONOS memory cell of Claim 66, wherein one of said control gates is a selected control gate and its underlying nitride region is a selected nitride region, and the other of said control gates is an unselected control gate and its underlying nitride region is an unselected nitride region, and wherein said bit line diffusion near said selected nitride region is a bit diffusion, and said bit line diffusion near said unselected nitride region is a source diffusion, wherein a program operation of said cell is performed by:

providing a high voltage on said unselected control gate to over-ride said unselected nitride region; and

varying a voltage on said selected control gate.

78. The MONOS memory cell of Claim 66 wherein said memory cell is one of many cells in a flash memory array that share a word line, and further comprising simultaneously programming several of said cells with different threshold levels by varying the voltage either of said control gate or said bit diffusion.

79. The MONOS memory cell of Claim 66, wherein an erase operation of a block of nitride regions is performed by:

providing a positive voltage to said bit line diffusions; and

providing a negative voltage to said control gates over said bit line diffusions.

80. The MONOS memory cell of Claim 66, wherein an erase operation of a block of nitride regions is performed by:

providing a negative voltage to said semiconductor substrate and to said bit line diffusions; and providing a positive voltage to said control gates.

81. A method of reading a MONOS memory cell, wherein the MONOS memory cell comprises:

a word gate on the surface of a semiconductor substrate;

sidewall control gates on sidewalls of said word gate, separated from said word gates by an insulating layer;

nitride regions within an ONO layer underlying said sidewall control gates wherein electron memory storage is performed within said nitride regions;

a polysilicon word line overlying and connecting said word gate with word gates in other said memory cells and overlying said sidewall control gates, separated from said sidewall control gates by an insulating layer; and

bit line diffusions within said semiconductor substrate adjacent to each of said sidewall control gates.

wherein one of said nitride regions is a selected nitride region, and the other of said nitride regions is an unselected nitride region, and wherein said bit line diffusion near said selected nitride region is a bit diffusion, and said bit line diffusion near said unselected nitride region is a source diffusion,

wherein a read operation of said cell is performed by:

over-riding said unselected nitride region;

providing a voltage on said word gate having a sum of the word gate threshold voltage, an overdrive voltage, and the voltage on said source diffusion;

providing a voltage on said control gate adjacent to said selected nitride region sufficient to allow for reading of the selected nitride region; and

reading said cell by measuring the voltage level on said bit diffusion.

82. A method of programming a MONOS memory cell, wherein said MONOS memory cell comprises:

a word gate on the surface of a semiconductor substrate;

sidewall control gates on sidewalls of said word gate, separated from said word gates by an insulating layer;

nitride regions within an ONO layer underlying said sidewall control gates wherein electron memory storage is performed within said nitride regions;

a polysilicon word line overlying and connecting said word gate with word gates in other said memory cells and overlying said sidewall control gates, separated from said sidewall control gates by an insulating layer; and

bit line diffusions within said semiconductor substrate adjacent to each of said sidewall control gates;

wherein one of said control gates is a selected control gate and its underlying nitride region is a selected nitride region, and the other of said control gates is an unselected control gate and its underlying nitride region is an unselected nitride region, and wherein said bit line diffusion near said selected nitride region is a bit diffusion, and said bit line diffusion near said unselected nitride region is a source diffusion,

wherein said method of programming the cell comprises the steps of:

providing a high voltage on said unselected control gate to over-ride said unselected nitride region; and

varying a voltage on said selected control gate.

83. A method of erasing a MONOS memory cell, wherein said MONOS memory cell comprises:

a word gate on the surface of a semiconductor substrate;

sidewall control gates on sidewalls of said word gate, separated from said word gates by an insulating layer;

nitride regions within an ONO layer underlying said sidewall control gates wherein electron memory storage is performed within said nitride regions;

a polysilicon word line overlying and connecting said word gate with word gates in other said memory cells and overlying said sidewall control gates, separated from said sidewall control gates by an insulating layer; and

bit line diffusions within said semiconductor substrate adjacent to each of said sidewall control gates;

wherein said method of erasing a block of said nitride regions comprises the steps of:

providing a positive voltage to said bit line diffusions; and

providing a negative voltage to said control gate over said bit line diffusions.

84. A flash memory device comprising:

word gates on the surface of a semiconductor substrate;

sidewall control gates on the sidewalls of said word gates separated from said word gates by an insulating layer;

bit line diffusions within said semiconductor substrate between two of said sidewall control gates; and

nitride charge regions underlying said sidewall control gates.

85. The device according to Claim 84 further comprising:

an insulating layer overlying said sidewall control gates; and

a word line overlying said control gates and connecting said word gates.

86. The MONOS memory call of Claim 84 wherein a channel length defined from an edge of said word gate to an edge of adjacent said bit diffusion region is between about 30 and 50 nm and whereby ballistic electron injection occurs.